

# Model Building on Selectivity of Gas Antisolvent Fractionation Method Using the Solubility Parameter

Máté Mihalovits<sup>1\*</sup>, Attila Horváth<sup>1</sup>, László Lőrincz<sup>1</sup>, Edit Székely<sup>1</sup>, Sándor Kemény<sup>1</sup>

<sup>1</sup> Department of Chemical and Environmental Process Engineering, Faculty of Chemical Technology and Biotechnology, Budapest University of Technology and Economics, H-1111 Budapest, Műegyetem rakpart 3., Hungary

\* Corresponding author, e-mail: [mihalovits@mail.bme.hu](mailto:mihalovits@mail.bme.hu)

Received: 15 July 2018, Accepted: 06 December 2018, Published online: 19 March 2019

## Abstract

Solubility parameters are widely used in the polymer industry and are often applied in the high pressure field as well as they give the possibility of combining the effects of all operational parameters on solubility in a single term. We demonstrate a statistical methodology to apply solubility parameters in constructing a model to describe antisolvent fractionation based chiral resolution, which is a complex process including a chemical equilibrium, precipitation and extraction as well. The solubility parameter used in this article, is the Hansen parameter. The evaluation of experimental results of resolution and crystallization of ibuprofen with (R)-phenylethylamine based on diastereomeric salt formation by gas antisolvent fractionation method was carried out. Two sets of experiments were performed, one with methanol as organic solvent in an undesigned experiment and one with ethanol in a designed experiment. The utilization of D-optimal design in order to decrease the necessary number of experiments and to overcome the problem of constrained design space was demonstrated. Linear models including dependence of pressure, temperature and the solubility parameter were appropriate to describe the selectivity of the GASF optical resolution method in both sets of experiments.

## Keywords

gas antisolvent precipitation, supercritical carbon dioxide, Hansen parameter, model building, D-optimal design

## 1 Introduction

Gas antisolvent precipitation (GAS) is one of the most promising innovative applications of supercritical fluids of the last decades. Its semicontinuous scaled up version might be also interesting for continuous pharmaceutical production [1]. During GAS processing the environmentally benign carbon dioxide is mixed intimately with an organic solvent containing the solute to be precipitated. Due to the instant and high oversaturation, small crystals or amorphous particles are formed with controllable mean particle sizes, narrow size distributions and sometimes controllable morphologies as well. Applications of and parameter effects on antisolvent precipitations were extensively reviewed [2-4]. While for typical GAS applications the goal is to completely precipitate the solute in a desired size, crystallization habit and morphology, gas antisolvent fractionation (GASF) is a combination of the precipitation and an extraction step [5, 6]. The GASF technique can be efficiently used in diastereomeric salt based optical resolutions [7-11] and purification of scalemic mixtures as well [7, 12, 13]. However, the

efficiency of GASF is influenced by various operational parameters. The most important ones are the pressure, the temperature, the concentration of the solutes, the solvent and the CO<sub>2</sub>. Furthermore, their effects are often non-linear. Our goal was to find a better way to optimize the optical resolutions with GASF than evaluating the effect of all these parameters individually.

The efficiency of the GASF optical resolution method can be characterized as the selectivity (*S*) of the process. Selectivity is calculated as the product of yield and diastereomeric purity. To describe selectivity, pressure (*p*), temperature (*T*) and the Hansen solubility parameter (HSP,  $\delta$ ) are taken into account. HSP is an attribute which describes how likely one material dissolves in another to form solution. Materials with similar HSP mix well with each other. Beside the nature of the components, HSP also depends on temperature and pressure. The formulas for calculating HSP of pure materials can be found in the literature [14] also for carbon dioxide and organic solvents [15] and their mixtures [16]. The parameter optimization difficulties of

supercritical antisolvent precipitation (or related) techniques have already been addressed using (Hansen) solubility parameters. However, only limited correlations were found [17–20].

Despite the fact that HSP itself is a function of the conditions of the operating system (pressure, temperature, quality of the components of the solvent mixture, ratio of the components of the solvent mixture), it cannot be expected that the selectivity could be described by this parameter only, because optical resolutions are not barely solubility controlled. That is the reason why we took into account temperature and pressure as well independently in our model. Evaluations of two experimental datasets are interpreted in this paper. The first one belongs to the antisolvent process performed with methanol as organic solvent. The second one includes data gained in the experiment performed with ethanol. Designed experiment was utilized in the experiments with ethanol, while in the case of methanol it was not utilized. The reason for lack of designed experiment in the case of methanol is that the experiment was being conducted without statistical support.

## 2 Methodology of model building

Two stages of model building are considered in this paper:

- model selection,
- residual analysis.

In the phase of model selection, the aim is to find a model which suitably describes and predicts the behavior of the observed system while it remains as simple as possible. Afterward a candidate model is selected, residual analysis is used to check the assumptions of regression method, to verify the adequacy of the selected model and to identify outlier data.

### 2.1 Model selection

There are two basic philosophies behind model selection. First, one would want the model to describe the given data as well as possible and has as high predictive accuracy as possible. This can be achieved by adding more and more parameters and different functions of the parameters (quadratic terms, cross-products, logarithms, etc.) to the model. On the other hand, one would want the model to be simple with as few terms as possible. The reason for the latter is that this way the model is easier to handle, to calculate with, and to understand the behavior of the observed system [21]. Also as Occam's razor (i.e. law of parsimony)

states, the simplest solution tends to be the correct one, therefore it is preferable to choose the simplest one.

There is no perfect model that could be selected in any situation. As Box and Draper [22] describe: "The most that can be expected from any model is that it can supply a useful approximation to reality: All models are wrong; some models are useful". It is the task of the experimenter to choose an appropriate model in view of the above-mentioned philosophies.

There are different statistical methods that can be used to choose the most suitable model [21]. In this paper, best subset regression procedure is being used. Other methods like stepwise regression and backward elimination could also be used, however they are more constrained and give less freedom in choosing from candidate models. The best subset method gives information about the goodness of fit of all possible models that can be built by the parameters. Afterwards the preferable model can be chosen by considering the goodness of fit (i.e. descriptive power) and the number of terms in the model. There are different indicators to describe the goodness of fit [23], in this paper the adjusted  $R^2$  is applied.

The higher the adjusted  $R^2$  the higher is the descriptive power of the model. However, adjusted  $R^2$  is a random variable, a higher  $R^2$  may be observed only by chance when there is no true difference in expected values. Two models with adjusted  $R^2$  relatively close to each other are worth to be compared based on the residual variances of the models. Significantly lower residual variance indicates better goodness of fit (significantly higher adjusted  $R^2$ ). The statistical comparison of variances is carried out with F-test.

A more sensitive test to compare models is based on the test of the extra sum of squares [24]. The extra sum of squares measures the reduction in the error sum of squares of a given model when terms are added to it. F-test is used to test the null hypothesis that the expected value of the extra error sum of squares is zero, i.e. adding these terms to the model does not increase the descriptive power of the model significantly. The test statistic is calculated by Eq. (1)

$$F_0 = \frac{(SS_q - SS_r)/(r - q)}{SS_r/(n - r - 1)}, \quad (1)$$

where  $SS_q$  is the error sum of squares of the basic model,  $q$  is the number of terms in that model,  $SS_r$  is the error sum of squares of the model with added terms in it,  $r$  is the total number of terms in that model and  $n$  is the number of data points. If the test statistic,  $F_0$  does not exceed the F-critical value at 5 % significance level with degrees of freedoms of

$(r - q)$  and  $(n - r - 1)$  the null hypothesis is accepted. Thus there is no significant difference between the two models.

## 2.2 Residual analysis

### 2.2.1 Checking assumptions

Assumptions of regression method have to be fulfilled otherwise the estimates of the regression parameters might be biased. In cases where the assumptions are not fulfilled, other statistical regression methods (e.g. Generalized Linear Regression, Nonparametric Regression) are to be used. Normality of the residuals with constant variance and independence of the residuals are those to be tested. Fulfilled assumptions are indicators of a proper model, therefore checking them also supports the verification of the model. The methods to check assumptions are not discussed in this paper, but demonstrated in the data evaluation section.

### 2.2.2 Outlier data detection

Identifying and removing outlier data is of importance, as they may distort the model estimation. A single point can be outlier in two ways: it can be outlier in the space of the dependent variable or in the space of the independent variables.

Plotting studentized residuals against predicted values [25] is a way to identify potential outlier data in the space of dependent variable. The points should scatter around zero approximately within a constant width band. Statistical software usually provides the studentized values thus they can be easily plotted. Usually values higher than 2.5 should be checked more closely as there is a great chance that these points are outliers.

Mahalanobis distance ( $D_i$ ) is used to identify data that are outlier in the space of the independent variables. The  $D_i$  values for the data are usually provided by statistical programs (calculations of  $D_i$  can be found for example in [26]). Plotting  $D_i$  against predicted values can make the examination easier by visualization. Points which are far from the rest may be considered outliers. Table for critical values for  $D_i$  can be found in [27].  $D_i$  higher than the critical value is considered outlier in the space of independent variables and should be removed from the dataset.

Cook's distance ( $C_i$ ) measures the combined influence on the case of being an outlier in the space of the dependent variable and the independent variables. Parameters of the fitted model significantly change when data with high  $C_i$  (formulas for calculations can be found in [28]) is omitted from the data set [29]. Cook suggests that data with  $C_i$  higher than 1 should be flagged as outlier and omitting the data should be considered [30].

## 3 Methodology of designed experiments

Design of experiments (DOE) is a statistical tool that constructs an experiment with least necessary experiments needed to be performed for creating a certain model. The structure of experiment designed by DOE can be made to be orthogonal which means that the parameters of the fitted model are estimated independently from each other. This results in a less uncertain model estimation than that of non-designed experiments [31].

Due to the fact that the solubility parameter itself depends on the other two controlled parameters (temperature and pressure) the settings of the experiments in a designed experiment is not straightforward. Three parameters define HSP: carbon-dioxide – organic solvent ratio, pressure and temperature. In order to tune HSP at any given set of pressure and temperature combination, the ratio of the solvent mixture is to be set. However, this ratio is constrained by the solubility of the ibuprofen and the solubility of the diastereomeric salt forms in the organic solvents. As experiments were designed with constant apparent concentrations of both the ibuprofen and of the salts in the GASF crystallizer, decreasing the amount of the organic solvent was limited by the solubility of the components in it. Increasing the amount of the organic solvent, which results in an increasing solubility of all components of interest, was limited as well since no fractionation could be possible without a diastereomeric salt precipitation. In practice due to the constraint, the HSP cannot be set to constant maximum and minimum values during the experiment as it would be desired in an orthogonal design. This means that full orthogonality cannot be achieved, approximate only. However, other advantages, such as smaller variance of estimates, can be gained over undesigned experiments by designing the experiment.

In our case there are 3 parameters, namely pressure, temperature and HSP. The expectation of the authors is that the selectivity of the antisolvent process can be described mostly by the linear effect of HSP, while temperature and pressure are less important factors and take part in the model only as correction terms. The effects that are evaluated are the linear effects of each parameter, the quadratic effects of temperature and pressure and the interaction of temperature and pressure. To estimate these effects a  $2 \times 3^2$  design would be desired, where HSP would be set at 2 levels while temperature and pressure would be set at 3 levels during the experiments. However, as it was discussed above, the setting of HSP is constrained, an orthogonal  $2 \times 3^2$  design cannot

be constructed. But designs with other advantages over undesigned experiment can be constructed. The frequently used D-optimal design [32] minimizes the volume of joint confidence region of estimated parameters of the model, i.e. minimizes the uncertainty of the estimates. TIBCO Statistica Software was used to create the D-optimal design for experiments in ethanol with Sequential algorithm [33].

## 4 Data evaluation

### 4.1 Experiment with methanol as the solvent

The dataset contains 34 data points that were obtained in these experiments. Exact experimental methodology can be found our previous paper [34]. The pressure, temperature and Hansen solubility parameter were varied in the experiments in a non-designed structure. The aim is to find an appropriate model to describe selectivity ( $S$ ) as function of temperature ( $T$ ), pressure ( $p$ ) and the Hansen parameter ( $\delta$ ). The data can be found in Table 1.

#### 4.1.1 Model selection

The following effects were considered in the candidate models: linear effects of temperature, pressure and Hansen parameter, quadratic effects of temperature and pressure and the interaction of temperature and pressure. The choice of the candidate effects are arbitrary and based on the preferences and expectations of the experimenters. If the data cannot be described appropriately with a model including these effects other effects should be considered to be taken into account.

The 10 best (based on adjusted  $R^2$ ) models obtained from best subset method can be seen in Table 2. Every row belongs to a certain model and the + signals mark those terms that are included in that model. Based on the discussion of model selection above, higher adjusted  $R^2$  and smaller number of effects are desired. Therefore #1 and #5 were chosen for further comparison.

The earlier was chosen because of its high adjusted  $R^2$  value and the latter was chosen because of its simplicity (and reasonably high  $R^2$  value). Model #1 has residual variance of 0.003276 with 29 degrees of freedom while model #5 has residual variance of 0.003584 with 30 degrees of freedom (not shown in the table). The two  $R^2$  values are relatively close to each other, the comparison is desirable. To compare the two models, the null hypothesis of the extra sum of square being zero is tested. The  $F_0$  is 3.82, calculated by Eq. (1). The critical F-value at 5 % significance level with degrees of freedom of 1 and 29 is 4.18.

**Table 1** Data of experiments with methanol

#	$p$ (MPa)	$T$ (°C)	$\delta$ (MPa <sup>0.5</sup> )	$S$ (-)
1	15	45	12.0	0.541
2	15	45	12.2	0.388
3	15	45	12.4	0.328
4	15	45	12.4	0.328
5	15	45	12.6	0.152
6	15	45	12.9	0.002
7	10	45	11.2	0.439
8	11	45	11.5	0.519
9	12	45	11.8	0.502
10	13	45	12.0	0.434
11	14	45	12.2	0.442
12	15	45	12.4	0.378
13	16	45	12.6	0.333
14	17	45	12.8	0.255
15	18	45	13.0	0.374
16	19	45	13.0	0.258
17	20	45	13.3	0.295
18	21	45	13.4	0.254
19	15	35	13.2	0.399
20	15	55	11.5	0.303
21	10	35	12.2	0.444
22	20	55	12.5	0.243
23	20	35	14.1	0.325
24	20	45	13.3	0.249
25	15	45	12.4	0.350
26	10	45	11.2	0.458
27	13	45	12.0	0.402
28	13	45	12.0	0.426
29	15	45	12.3	0.433
30	15	45	12.3	0.435
31	10	45	11.2	0.433
32	15	45	12.4	0.354
33	20	45	13.3	0.219
34	15	45	12.4	0.328

As  $F_0$  is smaller than the critical F-value, the null hypothesis is accepted, there is no significant difference in the descriptive power of the two models. This result supports the choice of the simpler model with only linear terms in it, namely  $p$ ,  $T$  and Hansen solubility parameter (model #5).

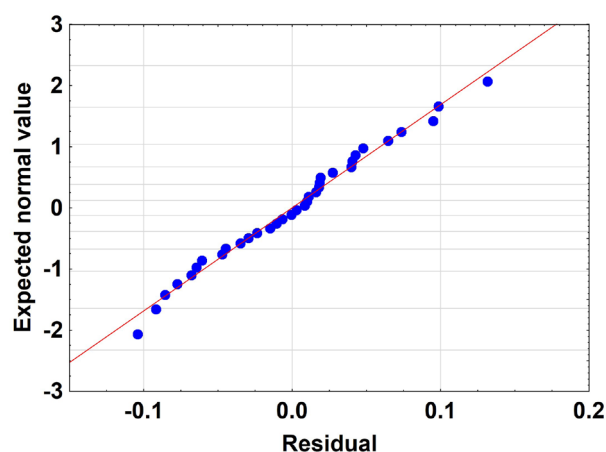
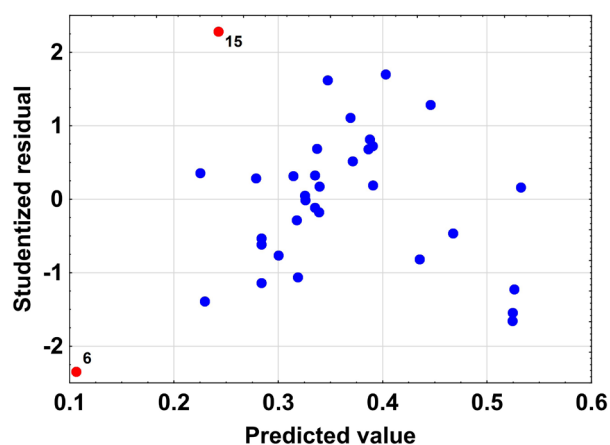
#### 4.1.2 Residual analysis

Residual analysis was carried out on the selected model in order to detect outlier data and to check assumptions of regression analysis and adequacy of the model.

**Table 2** Candidate models for further comparison

#	Adjusted $R^2$	Number of effects	$p$	$T$	$\delta$	$p * T$	$p^2$	$T^2$
1	0.73	4	+	+	+		+	
2	0.72	5	+	+	+	+	+	
3	0.72	5	+	+	+		+	+
4	0.71	6	+	+	+	+	+	+
5	0.70	3	+	+	+			
6	0.69	4	+	+	+	+		
7	0.69	4	+	+	+			+
8	0.68	4	+		+		+	+
9	0.68	5	+	+	+	+		+
10	0.68	3	+		+			+

Normality of residuals was examined with normal probability plot (Fig. 1). The data fit well to the red line representing the normal distribution and no deviation can be seen. Shapiro-Wilk's test [35] verifies the normality of residuals ( $p = 0.98$ ).

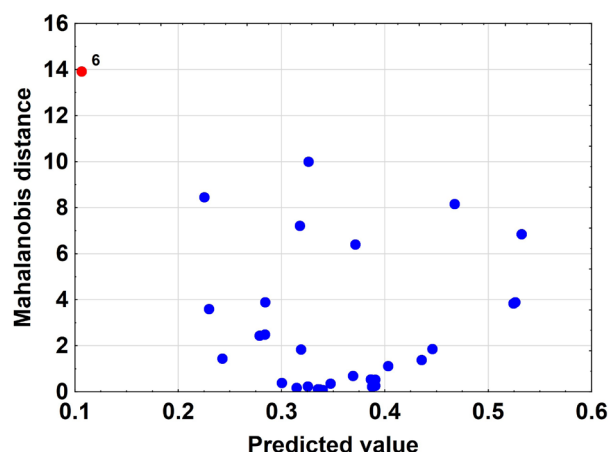
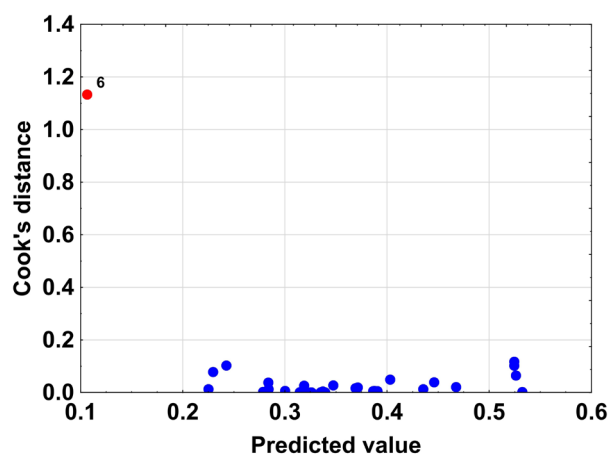

**Fig. 1** Normal probability plot of residuals

**Fig. 2** Studentized residuals against the predicted values.

In order to detect outlier data studentized residuals, Mahalanobis distances and Cook's distances were examined.

Fig. 2 shows the studentized residuals against the predicted values. Though the point #6 and #15 are a bit extreme compared to the rest, the values are lower than 2.5 and one has no reason to flag these points as outliers based on this plot. A potential anomaly can be seen in the plot as if the residuals had a curve trend. This pattern might suggest a need for quadratic term in the model. However, extended nature of the trend is not convincing, it should not necessarily give rise to concern.

Fig. 3 shows the Mahalanobis distances against the predicted values. Based on the table of critical values of Mahalanobis distances [27] the point #6 is outlier in the dimension of the independent variables at 5% significance level.

To affirm the outlier nature of this point, Cook's distances are examined. Fig. 4 shows the Cook's distances against the predicted values.


**Fig. 3** Mahalanobis distances against predicted values

**Fig. 4** Cook's distances against predicted values



The Cook's distance of point #6 is evidently higher than those of the other points and also higher than 1 which is a suggested limit in [30] to flag the point as outlier. Based on the above observations, point #6 is flagged as outlier and dropped from the dataset. It is not evident whether the point is outlier only in the space of the independent variables or both in the space of the independent variables and the dependent variable. Added experiences could be conducted with settings close to those of point #6 in order to test the outlier nature in the space of the dependent variable. Also these added experiences would reduce the extent of outlier nature in the space of independent variables of point #6.

As point #6 is dropped from the initial data set, the evaluation of the remaining data should be carried out from the beginning. The same model was chosen in the model selection phase as before, the assumptions of the analysis and the model adequacy were verified by residual analysis and no further outlier data were detected.

#### 4.1.3 Fitted model

The selectivity of the GASF optical resolution of ibuprofen from methanol can be described by the temperature, the pressure and the Hansen solubility parameter. The fitted model is presented in Eq. (2).

$$S = 5.20 + 0.0460p - 0.0312T - 0.334\delta, \quad (2)$$

where  $p$  is pressure (MPa),  $T$  is temperature ( $^{\circ}\text{C}$ ),  $\delta$  is Hansen solubility parameter ( $\text{MPa}^{0.5}$ ) and  $S$  is selectivity (-). The coefficient of HSP being much greater than the two other parameters, while the scales of parameters are within the same magnitude reinforces the expectations of the authors that HSP has the highest effect on the selectivity.

#### 4.2 Experiments with ethanol as the solvent

The second set of experiments was conducted in order to evaluate the process with ethanol as solvent in a designed experiment. Despite the fact that the response showed only linear correlation with the parameters in the experiment with methanol, a design was constructed that allows evaluation of quadratic effects of pressure and temperature as well. This experiment would contain 18 points ( $2 \times 3^2$ ) in an orthogonal design. The difficulty is that the design space of the experiments is constrained as it was discussed in Section 3. The settings of the Hansen parameter were the closest that was practically achievable to the settings of those in an orthogonal  $2 \times 3^2$  design.

Table 3 shows the settings of the experiments. The  $\delta^*$  column represents the settings as if the design space were not constrained, while the  $\delta$  column represents the HSP values that would be practically achievable. Also, the experimenters preferred to reduce the number of experiments to 14. D-optimal design method was used to construct a design with 14 points (chosen from among the practically achievable 18 points of the  $2 \times 3^2$  design) which allows evaluating quadratic effects of temperature and pressure as well as linear effects of the parameters and cross-product of temperature and pressure.

If it is found that considering these effects only, the descriptive power of the model is low, the experimenter should include other effects. However, to include other effects, it might be needed to complete the designed experiment with more experiments. The situation is special here, as the design space is restricted and the design is not orthogonal. It results in points (Table 3, the points with HSP around 13.0) that can be used to evaluate quadratic effect of HSP as well. The quadratic effect is not considered in the paper, as it was found to be the way the experimenters expected, non-significant.

The coloring in Table 3 marks the chosen points. Experiment #3, #8 and #17 were selected two times by the optimizing algorithm to enhance these settings in order to attain D-optimality.

**Table 3** Design of experiment

#	$p$ (MPa)	$T$ ( $^{\circ}\text{C}$ )	$\delta^*$ ( $\text{MPa}^{0.5}$ )	$\delta$ ( $\text{MPa}^{0.5}$ )
1	15	40	12	12.3
2	15	40	14	13.3
3	15	45	12	12.0
4	15	45	14	12.9
5	15	35	12	12.8
6	15	35	14	13.7
7	20	45	12	12.8
8	20	45	14	13.7
9	20	35	12	13.6
10	20	35	14	14.0
11	20	40	12	13.2
12	20	40	14	14.0
13	10	40	12	12.0
14	10	40	14	12.4
15	10	45	12	12.0
16	10	45	14	12.1
17	10	35	12	12.0
18	10	35	14	12.8

The experimental results can be found in Table 4. There are varying differences between the Hansen values defined in the designed experiment (Table 3,  $\delta$  column) and those are set in the experiments (Table 4,  $\delta$  column). The reason for that it is hard to set the parameter to an exact value, therefore what was achieved in the experiments those may differ from the ones defined in the design. The real values that were set during the experiments were recorded and the calculations were performed with them.

#### 4.2.1 Model selection

The simplest model which is linear in all the parameters is more than appropriate to describe the data, comparison is not needed with other models. The adjusted  $R^2$  is 0.99 while the residual variance is 0.000526. Note, that this residual standard variance is surprisingly smaller than the ones in the experiments with methanol as the solvent.

#### 4.2.2 Residual analysis

Residual analysis was carried out on the selected model, the details are not shown here. The assumptions of the regression analysis are verified, there was no sign of the model being inadequate. Also no data was found to be outlier.

#### 4.2.3 Fitted model

The selectivity of the antisolvent fragmentation of ibuprofen from ethanol can be described by the temperature, the pressure and the Hansen solubility parameter. The model fitted is given in Eq. (3)

$$S = 6.06 + 0.0643p - 0.0389T - 0.400\delta, \quad (3)$$

where  $p$  is pressure (MPa),  $T$  is temperature ( $^{\circ}\text{C}$ ),  $\delta$  is Hansen solubility parameter ( $\text{MPa}^{0.5}$ ) and  $S$  is selectivity (-). It can be stated here as well that HSP has the highest effect on the response,  $S$ .

#### 4.3 Comparison of the estimates of the two models

It can be seen in Table 5, that the estimates of the parameters in the two models are quite similar. Further analysis could be carried out to test whether the two datasets can be described by a single model, however it is not in the focus in this paper. The standard errors of the estimates are function of the residual variance, the number of data point, and the structure of the design of the experiment. Smaller residual variance, greater number of data points and designed nature of the experiments are decreasing the standard errors of estimates.

**Table 4** Data of the experiments with ethanol

#	$p$ (MPa)	$T$ ( $^{\circ}\text{C}$ )	$\delta$ ( $\text{MPa}^{0.5}$ )	$S$ (-)
1	10	45	12.237	0.033
2	10	40	12.700	0.107
3	15	40	13.575	0.024
4	20	45	13.987	0.012
5	15	35	13.949	0.078
6	20	45	14.028	0.000
7	15	35	13.926	0.088
8	10	35	12.017	0.559
9	15	45	12.068	0.466
10	10	35	12.016	0.517
11	15	45	12.083	0.416
12	15	40	12.538	0.477
13	20	35	13.819	0.456
14	20	40	13.427	0.416

**Table 5** Parameter estimates and its standard errors

	Methanol		Ethanol	
	Estimate	Standard error	Estimate	Standard error
Intercept	5.20	1.23	6.06	0.165
$p$	0.0460	0.0187	0.0643	0.00266
$T$	-0.0312	0.00823	-0.0389	0.00166
$\delta$	-0.334	0.0932	-0.400	0.0120

Despite the fact that the designed ethanol experiment contained half the number of data compared to the undesigned methanol experiment, the standard errors of the parameters of the ethanol model are much smaller. This is the effects of the residual variance being much smaller in the ethanol experiments and the designed nature of the experiment. Moreover, the probable reason for the smaller residual variance is the D-optimal property of the design. The 14 experiments with ethanol were performed consecutively, in relatively short time, in a designed structure. Contrary, the previous experiments with methanol were done less consecutively in longer time scale without utilization of DOE. The discontinuity of the experiments gave opportunity to random effects (effect of the day for example) to arise and increase the variance of the process significantly, making the residual variances greater.

#### 5 Conclusion

Evaluation and model building utilizing statistical methods were interpreted in the paper for GASF experiments with methanol and ethanol as organic-solvents. A proper model was found to describe the selectivity of the GASF

method by operating parameters, namely pressure, temperature and Hansen solubility parameter. The model shows that HSP has the highest effect on selectivity. We have no knowledge about other study in the literature which focuses on the modelling of the GASF method process especially not one that includes the Hansen solubility parameter. This quite simple model can easily be used to predict the selectivity of the process at given operating conditions. Also it can be used to find the optimal operating parameters where the selectivity is the highest. Furthermore it was demonstrated that optimal design is a remarkable tool in researches with constrained design space like the supercritical antisolvent method. The design

does not only reduce the number of experiments required to build a certain model but also reduces the uncertainty of the model estimates.

### Acknowledgement

This work is related to the National Research, Development and Innovation Fund of Hungary in the frame of FIEK\_16-1- 2016-0007 (Higher Education and Industrial Cooperation Center). The authors would like to acknowledge the financial support of Gedeon Richter Talentum Foundation (Ph.D. scholarship to M.M.). The authors thank the fruitful discussions and support of Csaba Dezső András, Éva Pusztai and Kinga Komka.

### References

- [1] Kankala, R. K., Zhang, Y. S., Wang, S.-B., Lee, C.-H., Chen, A.-Z. "Supercritical Fluid Technology: An Emphasis on Drug Delivery and Related Biomedical Applications", *Advanced Healthcare Materials*, 6(16), article ID: 1700433, 2017. <https://doi.org/10.1002/adhm.201700433>
- [2] Jung, J., Perrut, M. "Particle design using supercritical fluids: Literature and patent survey", *The Journal of Supercritical Fluids*, 20(3), pp. 179–219, 2001. [https://doi.org/10.1016/S0896-8446\(01\)00064-X](https://doi.org/10.1016/S0896-8446(01)00064-X)
- [3] Reverchon, E., De Marco, I. "Mechanisms controlling supercritical antisolvent precipitate morphology", *Chemical Engineering Journal*, 169(1-3), pp. 358–370, 2011. <https://doi.org/10.1016/j.cej.2011.02.064>
- [4] Knez, Ž., Hrnčič, M. K., Škerget, M. "Particle Formation and Product Formulation Using Supercritical Fluids", *Annual Review of Chemical and Biomolecular Engineering*, 6(1), pp. 379–407, 2015. <https://doi.org/10.1146/annurev-chembioeng-061114-123317>
- [5] Marqués, J. L., Porta, G. D., Reverchon, E., Renuncio, J. A. R., Mainar, A. M. "Supercritical antisolvent extraction of antioxidants from grape seeds after vinification", *The Journal of Supercritical Fluids*, 82, pp. 238–243, 2013. <https://doi.org/10.1016/J.SUPFLU.2013.07.005>
- [6] Sánchez-Camargo, A. P., Mendiola, J. A., Valdés, A., Castro-Puyana, M., García-Cañas, V., Cifuentes, A., Herrero, M., Ibáñez, E. "Supercritical antisolvent fractionation of rosemary extracts obtained by pressurized liquid extraction to enhance their antiproliferative activity", *The Journal of Supercritical Fluids*, 107, pp. 581–589, 2016. <https://doi.org/10.1016/J.SUPFLU.2015.07.019>
- [7] Kordikowski, A., York, P., Latham, D. "Resolution of ephedrine in supercritical CO<sub>2</sub>: A novel technique for the separation of chiral drugs", *Journal of Pharmaceutical Sciences*, 88(8), pp. 786–791, 1999. <https://doi.org/10.1021/js980459f>
- [8] Martín, A., Cocero, M. J. "Separation of enantiomers by diastereomeric salt formation and precipitation in supercritical carbon dioxide. Application to the resolution of mandelic acid", *The Journal of Supercritical Fluids*, 40(1), pp. 67–73, 2007. <https://doi.org/10.1016/j.supflu.2006.03.018>
- [9] Körösi, M., Sedon, A., Komka, K., Sohajda, T., Székely, E. "Gas Antisolvent Fractionation: a New Method to Obtain Enantiopure Compounds, a Case Study on Mandelic Acid", *Periodica Polytechnica Chemical Engineering*, 63(1), pp. 130–137, 2019. <https://doi.org/10.3311/PPCh.11741>
- [10] Bánsághi, G., Lőrincz, L., Szilágyi, I. M., Madarász, J., Székely, E. "Crystallization and Resolution of *cis*-Permethric Acid with Carbon Dioxide Antisolvent", *Chemical Engineering and Technology, Special Issue: Industrial Crystallization*, 37(8), pp. 1417–1421, 2014. <https://doi.org/10.1002/ceat.201300718>
- [11] Zodge, A., Körösi, M., Tárkányi, M., Madarász, J., Szilágyi, I. M., Sohajda, T., Székely, E. "Gas Antisolvent Approach for the Precipitation of  $\alpha$ -Methoxyphenylacetic Acid – (*R*)-1-Cyclohexylethylamine Diastereomeric Salt", *Chemical and Biochemical Engineering Quarterly*, 31(3), pp. 335–341, 2017. <https://doi.org/10.15255/CABEQ.2016.1023>
- [12] Körösi, M., Madarász, J., Sohajda, T., Székely, E. "A fast, new method to enhance the enantiomeric purity of non-racemic mixtures: self-disproportionation of enantiomers in the gas antisolvent fractionation of chlorine-substituted mandelic acid derivatives", *Tetrahedron: Asymmetry*, 28(11), pp. 1568–1572, 2017. <https://doi.org/10.1016/J.TETASY.2017.09.005>
- [13] Körösi, M., Madarász, J., Sohajda, T., Székely, E. "Fast further purification of diastereomeric salts of a nonracemic acid by gas antisolvent fractionation", *Chirality*, 29(10), pp. 610–615, 2017. <https://doi.org/10.1002/chir.22730>
- [14] Williams, L. L., Rubin, J. B., Edwards, H. W. "Calculation of Hansen Solubility Parameter Values for a Range of Pressure and Temperature Conditions, Including the Supercritical Fluid Region", *Industrial & Engineering Chemistry Research*, 43(16), pp. 4967–4972, 2004. <https://doi.org/10.1021/ie0497543>
- [15] Hansen, C. M. "Hansen Solubility Parameters: A User's Handbook", 2nd ed., CRC Press, Boca Raton, USA, 2007. <https://doi.org/10.1201/9781420006834>
- [16] Machida, H., Takesue, M., Smith Jr., R. L. "Green chemical processes with supercritical fluids: Properties, materials, separations and energy", *The Journal of Supercritical Fluids*, 60, pp. 2–15, 2011. <https://doi.org/10.1016/j.supflu.2011.04.016>



- [17] Park, S.-J., Jeon, S.-Y., Yeo, S.-D. "Recrystallization of a Pharmaceutical Compound Using Liquid and Supercritical Antisolvents", *Industrial & Engineering Chemistry Research*, 45(7), pp. 2287–2293, 2006.  
<https://doi.org/10.1021/ie0510775>
- [18] Pereira, V. J., Matos, R. L., Cardoso, S. G., Soares, R. O., Santana, G. L., Costa, G. M. N., Vieira de Melo, S. A. B. "A new approach to select solvents and operating conditions for supercritical anti-solvent precipitation processes by using solubility parameter and group contribution methods", *The Journal of Supercritical Fluids*, 81, pp. 128–146, 2013.  
<https://doi.org/10.1016/j.supflu.2013.05.010>
- [19] Seo, B., Kim, T., Park, H. J., Kim, J.-Y., Lee, K. D., Lee, J. M., Lee, Y.-W. "Extension of the Hansen solubility parameter concept to the micronization of cyclotrimethylenetrinitramine crystals by supercritical anti-solvent process", *The Journal of Supercritical Fluids*, 111, pp. 112–120, 2016.  
<https://doi.org/10.1016/j.supflu.2016.01.015>
- [20] Yoon, T. J., Son, W.-S., Park, H. J., Seo, B., Kim, T., Lee, Y.-W. "Tetracycline nanoparticles precipitation using supercritical and liquid CO<sub>2</sub> as antisolvents", *The Journal of Supercritical Fluids*, 107, pp. 51–60, 2016.  
<https://doi.org/10.1016/j.supflu.2015.08.014>
- [21] Draper, N. R., Smith, H. "Applied Regression Analysis", 3rd ed., Wiley, Somerset, NJ, USA, 2014.  
<https://doi.org/10.1002/9781118625590>
- [22] Box, G. E. P., Draper, N. R. "Empirical Model-Building and Response Surfaces", 1st ed., Wiley, New York, NY, USA, 1987.
- [23] Ryan, T. P. "Modern Regression Methods", 1st ed., Wiley, New York, NY, USA, 1997.
- [24] Kutner, M. H., Nachtsheim, C. J., Neter, J., Li, W. "Applied Linear Statistical Models", 5th ed., McGraw-Hill/Irwin, Boston, USA, 2005.
- [25] Weisberg, S. "Applied Linear Regression", 3rd ed., Wiley, Hoboken, New Jersey, USA, 2005.  
<https://doi.org/10.1002/0471704091>
- [26] Stevens, J. P. "Outliers and influential data points in regression analysis", *Psychological Bulletin*, 95(2), pp. 334–344, 1984.  
<https://doi.org/10.1037/0033-2909.95.2.334>
- [27] Barnett, V., Lewis, T. "Outliers in Statistical Data", 3rd ed., Wiley, Chichester, USA, 1994.
- [28] Cook, R. D. "Detection of Influential Observation in Linear Regression", *Technometrics*, 19(1), pp. 15–18, 1977.  
<https://doi.org/10.2307/1268249>
- [29] Cook, R. D. "Influential Observations in Linear Regression", *Journal of the American Statistical Association*, 74(365), pp. 169–174, 1979.  
<https://doi.org/10.2307/2286747>
- [30] Cook, R. D., Weisberg, S. "Residuals and Influence in Regression", Chapman and Hall, New York, NY, USA, 1982.
- [31] Box, G. E. P., Hunter, J. S., Hunter, W. G. "Statistics for Experimenters", 2nd ed., Wiley, Chichester, USA, 2005.
- [32] Myers, R. H., Montgomery, D. C., Anderson-Cook, C. M. "Response Surface Methodology: Process and Product Optimization Using Designed Experiments", 4th ed., Wiley, Hoboken, NJ, USA, 2016.
- [33] Dykstra Jr., O. "The Augmentation of Experimental Data to Maximize |X'X|", *Technometrics*, 13(3), pp. 682–688, 1971.  
<https://doi.org/10.2307/1267180>
- [34] Lőrincz, L., Bánsághi, G., Zsembéri, M., de Simón Brezmes, S., Szilágyi, I. M., Madarász, J., Sohajda, T., Székely, E. "Diastereomeric salt precipitation based resolution of ibuprofen by gas antisolvent method", *The Journal of Supercritical Fluids*, 118, pp. 48–53, 2016.  
<https://doi.org/10.1016/j.supflu.2016.07.021>
- [35] Shapiro, S. S., Wilk, M. B. "An analysis of variance test for normality (complete samples)", *Biometrika*, 52 (3-4), pp. 591–611, 1965.  
<https://doi.org/10.1093/biomet/52.3-4.591>